

WHAT IS CLAIMED IS:

1 1. A method for collecting data from among a plurality of data sites, each
2 data site having an associated data store, the method comprising:
3 providing each data site with a corresponding extraction routine;
4 for each data site, processing data contained in its associated data store in
5 accordance with its corresponding extraction routine to produce first data, the corresponding
6 extraction routine configured to store the first data in a storage location at the data site if the
7 processing produces first data;
8 collecting second data from those data sites for which their corresponding
9 extraction routines produced first data, the second data being based on the first data; and
10 loading all of the second data into a database.

1 2. The method of claim 1 wherein the step of collecting includes
2 communicating the first data to a central site in accordance with a remote copy operation,
3 receiving at the central site the first data as mirrored data, and transforming the mirrored data
4 to produce the second data, the second data being stored at the central site.

1 3. The method of claim 2 wherein the step of collecting comprises, at
2 each data site that has first data, performing a data backup procedure on the first data,
3 wherein the central site is viewed as a backup site.

1 4. The method of claim 2 wherein the step of communicating the first
2 data includes signaling the central site at time when the remote copy operation is complete,
3 the step of transforming the mirrored data being initiated in response to the central site being
4 signaled.

1 5. The method of claim 1 wherein the step of collecting includes:
2 transforming the first data to produce the second data, the second data being stored at the data
3 site; communicating the second data to a central site in accordance with a remote copy
4 operation, and receiving at the central site the second data as mirrored data.

1 6. The method of claim 1 further comprising providing each data site
2 with a corresponding remote copy program.

1 7. The method of claim 1 further including providing each data site with a
2 corresponding transformation routine, and for each data site having first data processing the
3 first data by executing its corresponding transformation routine to produce the second data.

1 8. The method of claim 1 wherein the step of providing a corresponding
2 extraction routine comprises, for each data site:
3 receiving a specification information which is descriptive of data stored in a
4 data store associated with the data site;
5 producing the extraction routine based on the specification information; and
6 communicating the extraction routine to the data site.

1 9. The method of claim 1 wherein the step of providing a corresponding
2 extraction routine comprises, for each data site:
3 receiving a first specification information which is descriptive of data stored in
4 its associated data store;
5 producing a second specification information which is descriptive of the
6 extraction routine;
7 communicating the second specification information to the data site; and
8 producing, at the data site, the extraction routine based on the second
9 specification information.

1 10. The method of claim 9 wherein the second specification information is
2 source code.

1 11. A data collection system comprising a plurality of remote data sites
2 and a data collection site, the remote data sites and the data collection site configured to
3 operate according to the method of claim 1.

1 12. A data collection system comprising:
2 a central data site comprising at least one host processor;
3 a host storage system operatively coupled to the at least one host processor;
4 at least one remote data site comprising at least one remote processor; and
5 a remote storage system operatively coupled to the at least one remote
6 processor,
7 the at least one host processor having program generating code configured to:

8 obtain storage related parameters from the at least one remote data site;
9 generate first interim volume managing code;
10 generate second interim volume managing code;
11 generate a data extraction routine based on the storage related
12 parameters;
13 generate remote copy control code based on the storage related
14 parameters; and
15 transfer the second interim volume managing code, the data extraction
16 routine, and the remote copy control code to the at least one remote data site,
17 the at least one host processor further having program manager code
18 configured to initiate processing of the second interim volume managing code, the data
19 extraction routine, and the remote copy control code,
20 the data extraction routine configured to produce extracted data,
21 the remote copy control code configured to perform a data duplication
22 operation of the extracted data, wherein the central data site serves as the duplication site for
23 the extracted data,
24 the first interim volume managing code configured to allocate a host interim
25 volume in the host storage system for storing at least some of the extracted data,
26 the second interim volume managing code configured to allocate a remote
27 interim volume in the remote storage system for storing the extracted data.

1 13. The system of claim 12 wherein the data duplication operation is a
2 remote copy operation.

1 14. The system of claim 12 wherein the at least some of the extracted data
2 is copied to the central data site as the result of a data mirroring operation.

1 15. The system of claim 12 wherein data extraction routine includes a size
2 calculation component to produce a data size metric indicative of the size of the extracted
3 data, the data size metric being communicated to the first interim volume managing code and
4 to the second interim volume managing code, wherein the host interim volume is allocated
5 based at least on the data size metric and the remote interim volume is allocated based at least
6 on the data size metric.

1 16. The system of claim 15 wherein the host interim volume has a flag
2 associated with the remote data site, the remote copy control code further configured to set
3 the flag to a first value when performing the backup operation and to set the flag to a second
4 value when the backup operation has completed.

1 17. A method for collecting data from among a plurality of remote data
2 sites comprising:
3 producing extraction routines, each extraction routine corresponding to a
4 remote data site, each extraction routine configured to process data stored at its corresponding
5 remote data site, the extraction routine further configured to store any data produced thereby
6 at its corresponding remote data site;
7 communicating each extraction routine to its corresponding remote data site;
8 initiating execution of each extraction routine to process the data stored at its
9 corresponding remote data site wherein if extracted data is produced then it is stored at its
10 corresponding remote data site, each extraction routine being executed at its corresponding
11 remote data site;
12 receiving processed data from each remote data site that produced extracted
13 data, the processed data being based on the extracted data; and
14 incorporating the processed data into a collection of information.

1 18. The method of claim 17 wherein the step of incorporating comprises
2 transforming the processed data to produce transformed data and loading the transformed
3 data into a database.

1 19. The method of claim 17 further including producing transformation
2 routines, wherein the step of transforming the processed data includes applying the
3 transformation routines to the processed data.

1 20. The method of claim 19 wherein receiving processed data includes
2 detecting an indication that the processed data from a remote data site has been
3 communicated therefrom and wherein transforming the processed data is performed in
4 response to the indication.

1 21. The method of claim 17 wherein the collection of information is stored
2 at a collection site and the extraction routines are produced at the collection site.

1 22. The method of claim 17 further comprising:
2 producing transformation routines, each transformation routine corresponding
3 to one of the remote data sites, the transformation routine further configured to store any data
4 produced thereby at its corresponding remote data site; and
5 communicating each transformation routine to its corresponding remote data
6 site,
7 each transformation routine configured to process the extracted data stored at
8 its corresponding remote data site to produce transformed data,
9 the processed data being representative of the transformed data.

1 23. The method of claim 22 wherein the collection of information is stored
2 at a collection site and wherein the transformation routines are produced at a collection site.

1 24. The method of claim 17 further comprising producing a remote copy
2 control routine for each of the remote data sites and communicating each remote copy control
3 routine to its corresponding remote data site.

1 25. The method of claim 17 further comprising:
2 creating a storage volume for storing the processed data;
3 receiving size information from one or more of the remote data sites, the size
4 information representative of the size of its corresponding extracted data; and
5 allocating space for the storage volume based on the size information.

1 26. An ETL (extraction, transformation, loading) tool configured to
2 perform the method of claim 17.

1 27. A data collection system comprising:
2 a data processing component;
3 a data storage component; and
4 computer program code configured to execute on the data processing
5 component, the computer program code configured to perform steps of:
6 receiving first information which specify extraction routines, each
7 extraction routine corresponding to a remote data site, each extraction routine
8 configured to process data stored at its corresponding remote data site, the extraction

9 routine further configured to store any data produced thereby at its corresponding
10 remote data site;
11 communicating to each of the remote data sites a corresponding second
12 information, the second information representative of the extraction routine
13 corresponding to the remote data site, whereby an extraction routine is provided at
14 each of the remote data sites; and
15 initiating execution of each extraction routine to process the data stored
16 at its corresponding remote data site wherein if extracted data is produced then it is
17 stored at its corresponding remote data site, each extraction routine being executed at
18 its corresponding remote data site,
19 a portion of the data storage component being configured to receive processed
20 data from each remote data site that produced extracted data, the processed data being
21 received as mirrored data resulting from a data mirroring operation performed at the remote
22 data sites, the processed data being based on the extracted data,
23 the computer program code further configured to perform a step of
24 incorporating the mirrored data into a database.

1 28. The system of claim 27 wherein the computer program code includes
2 an ETL (extraction, transformation, loading) tool, the ETL tool configured to produce a
3 plurality of extraction routines as the second information.

1 29. The system of claim 27 wherein the step of incorporating the mirrored
2 data includes transforming the mirrored data to produce transformed data and loading the
3 transformed data into the database.

1 30. The system of claim 27 wherein the computer program code is further
2 configured to:

3 produce transformation routines, each transformation routine corresponding to
4 one of the remote data sites, the transformation routine further configured to store any data
5 produced thereby at its corresponding remote data site; and

6 communicate each transformation routine to its corresponding remote data
7 site,

8 each transformation routine configured to process the extracted data stored at
9 its corresponding remote data site to produce transformed data,

10 the processed data being representative of the transformed data.

1 31. A method for processing data at a remote site to produce data suitable
2 for integration at a data collection site comprising:
3 obtaining an extraction routine at a remote site, the extraction routine suitable
4 for extracting data from a data store at the remote site, the extraction routine configured to
5 execute on a data processor at the remote site;
6 detecting a first signal and in response thereto, performing a data extraction
7 operation by applying the extraction routine to the data to produce extracted data;
8 storing a first representation of the extracted data at the remote site; and
9 performing a backup operation of the first representation of the extracted data
10 wherein the data is backed up to the data collection site.

1 32. The method of claim 31 wherein the first representation of the
2 extracted data is the extracted data itself, wherein a subsequent transformation operation and
3 a loading operation are performed at the collection site.

1 33. The method of claim 31 further comprising determining a size of the
2 data to be extracted and creating a logical volume having a storage capacity based on the size
3 of the data to be extracted, wherein the first representation of the extracted data is stored on
4 the logical volume.

1 34. The method of claim 33 further comprising determining a size of the
2 logical volume and communicating the size of the logical volume to the collection site, prior
3 to the step of performing a backup operation, wherein the collection site can allocate storage
4 thereat based on the size of the logical volume in order to store the first representation o of
5 the extracted data.

1 35. The method of claim 31 further comprising receiving a transformation
2 routine and applying the transformation routine to the extracted data to produce transformed
3 data, the first representation comprising the transformed data, wherein the transformed data is
4 backed up to the collection site.

1 36. The method of claim 31 wherein the step of obtaining an extraction
2 routine comprises receiving the extraction routine in executable form.

1 37. The method of claim 36 wherein the extraction routine is received
2 from the collection site.

1 38. The method of claim 31 wherein the step of obtaining an extraction
2 routine comprises receiving a representation of the extraction routine and generating an
3 executable form of the extraction routine from the representation.

1 39. A data storage facility configured to produce data suitable for
2 integration at a data collection site comprising:
3 a data processing component;
4 a data storage component; and
5 computer program code configured to perform steps of:
6 obtaining an extraction routine, the extraction routine suitable for
7 extracting first data from the data storage component, the extraction routine
8 configured to execute on the data processing component;
9 detecting a first signal and in response thereto, performing a data
10 extraction operation by applying the extraction routine to the first data to produce
11 extracted data; and
12 storing a first representation of the extracted data on the data storage
13 component,
14 the data storage component being configured to perform a backup operation of
15 the first representation of the extracted data wherein the data is backed up to the data
16 collection site.

1 40. The data storage facility of claim 39 wherein the computer program
2 code is further configured to receive a representation of the extraction routine and to generate
3 the extraction routine based on the representation, thereby obtaining the extraction routine.

1 41. The data storage facility of claim 39 wherein the data storage
2 component is further configured to perform a backup operation upon detecting an indication,
3 the indication being produced at the collection site.

1 42. A system for collecting data from among a plurality of data sites
2 comprising:

3 means for receiving first information specifying a data extraction routine,
4 wherein a plurality of data extraction routines can be produced, each data extraction routine
5 corresponding to a remote data site, each data extraction routine being configured based on
6 data contained at its corresponding remote data site;

7 means for communicating second information to each of the remote data sites,
8 the second information representative of its corresponding data extraction routine, wherein
9 each data extraction routine is executed at its corresponding remote data site;

10 means for signaling at least one of the data extraction routines to process data
11 at its corresponding remote data site to produce extracted data, wherein the at least one each
12 of the data extraction routines is configured to store extracted data in a data store at its
13 corresponding remote data site;

14 means for receiving extracted data from each of the remote data sites as
15 mirrored data;

16 means for transforming mirrored data to produce transformed data; and

17 means for loading transformed data into a data store.

1 43. The system of claim 42 wherein the second information constitutes the
2 data extraction routine.

1 44. The system of claim 42 wherein the second information comprises a
2 first representation of the data extraction routine, wherein the corresponding remote data site
3 can produce the data extraction routine based on the first representation.

1 45. The system of claim 42 further including means for producing a
2 remote copy control routine for each of the remote data sites and means for communicating
3 the remote copy control routines to the remote data sites.

1 46. A data storage facility configured to produce data suitable for
2 integration at a data collection site comprising:

3 means for obtaining an extraction routine at the remote site, the extraction
4 routine suitable for extracting data from a data store at the remote site, the extraction routine
5 configured to execute on a data processor at the remote site;

6 means for detecting a first signal and in response thereto for performing a data
7 extraction operation by applying the extraction routine to the data to produce extracted data;

8 means for storing a first representation of the extracted data at the remote site;
9 and
10 means for performing a backup operation of the first representation of the
11 extracted data wherein the data is backed up to the data collection site.

1 47. The data storage facility of claim 46 further including means for
2 obtaining a transformation routine at the remote site, means for applying the transformation
3 routine to the extracted data to produce transformed data, the first representation of the
4 extracted data being the transformed data.